Passive Acoustic Monitoring for the Detection and Identification of Marine Mammals

Marie A. Roch
5500 Campanile Drive
Department of Computer Science
San Diego, CA 92182-7720
phone: (619) 594-5830 fax: (619) 594-6746 e-mail: marie.roch@sdsu.edu

Grant Number: N000140811199

LONG-TERM GOALS

This project is intended to advance the state of passive acoustic monitoring. Improved methods of identifying cetaceans are developed in order to contribute to the Navy's mitigation efforts.

APPROACH

This project is a multi-pronged study to advance the state of the field in three areas. The development of automated auditory scene analysis for delphinid tonal calls will permit subsequent work by this investigator or others to exploit the use of whistles for classification and localization. Our approach is to dynamically build hypothesis graphs using phase-frequency representations of the signal. In parallel to this effort, two modeling techniques are being pursued to improve existing passive acoustic monitoring capabilities based on echolocation clicks of odontocetes. The first of these examines the use of a universal background model as proposed by Reynolds et al. (2000) for human speaker verification tasks. Reynolds' problem, which is similar in nature to ours, is how can one reject observations from a speaker (or dolphin species) for which there is no data to create a model. We adapt his idea of a universal background model by training a generalized odontocete model using the data of a number of species. This model is not specific to any one species. Using Bayesian learning, training data from a specific species adapts the parameters of the generalized model, thus serving as a foil against vocalizations that sound similar to one of our species. The second approach for echolocation clicks exploits recent machine learning work on submanifold learning (Dasgupta and Freund, 2007; Dasgupta and Freund, 2008; Freund et al., 2007). In order to detect and classify odontocetes, features, or poignant characteristics of their signals, must be extracted from the audio signal. As the underlying process of sound generation cannot be measured directly, nor is it well understood, classification techniques must attempt to infer information about the producer of the signal (e.g. species) through a typically higher-order set of features. Submanifold learners focus on learning a subspace of the high-order feature space that can be more conducive to providing robust classification.

WORK COMPLETED

The majority of the whistle extraction system has been implemented and we are completing scoring tools to evaluate the system. A framework for the universal background model detection system has

Report Documentation Page					*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE **2009**		2. REPORT TYPE			3. DATES COVERED **00-00-2009 to 00-00-2009**
4. TITLE AND SUBTITLE **Passive Acoustic Monitoring For The Detection And Identification Of Marine Mammals**					5a. CONTRACT NUMBER
					5b. GRANT NUMBER
					5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)					5d. PROJECT NUMBER
					5e. TASK NUMBER
					5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Computer Science,5500 Campanile Drive,San Diego,CA,92182**					8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					10. SPONSOR/MONITOR'S ACRONYM(S)
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT **This project is intended to advance the state of passive acoustic monitoring. Improved methods of identifying cetaceans are developed in order to contribute to the Navy?s mitigation efforts.**					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

been completed and evaluated using five species of delphinids from the Southern California Bight. We have created a framework for using the random projection tree submanifold learner and have extended an implementation of the algorithm provided by Freund and Dasgupta to provide pruning capabilities, a necessary component for tree-based classifiers which have a tendency to overlearn when not pruned.

RESULTS

Preliminary qualitative evaluations of the whistle contour extractor have been completed on the following species: bottlenose dolphins (*Tursiops truncatus*), melon-headed dolphins (*Peponocephela electra*), and long-beaked common dolphins (*Delphinus capensis*) and presented at conferences. Figure 1 shows a spectrogram containing many whistles and clicks and shows along with the detected whistles. Annotation is under way for ground truth information, although a recent conversation with Shannon Rankin (NOAA/NMFS) may open a better path to verification which will be investigated in the coming weeks. Informal tests show that the current methods are effective on signals with significant acoustic clutter in the auditory scene with the exception of clutter due to burst pulses which is a topic of continued research.

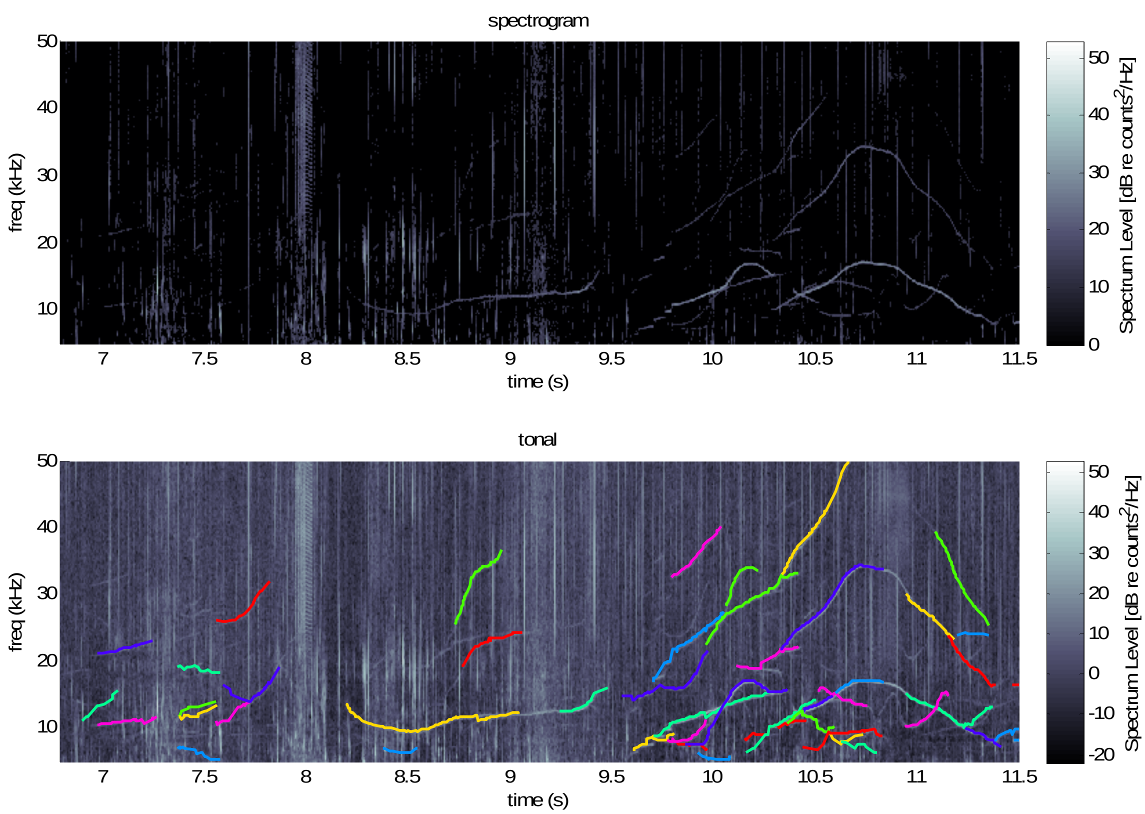
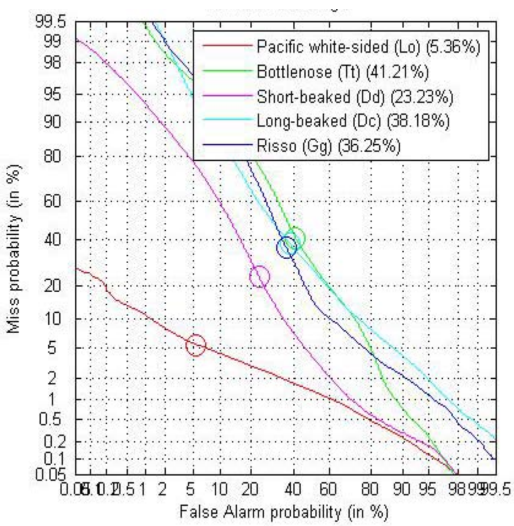


Figure 1 (color online) - Whistles extracted from long beaked common dolphin (Delphinus capensis) whistles with a threshold of 10 dB re counts2/Hz. The upper drawing shows a thresholded spectrogram with all time × frequency bins under 10 dB set to 0 and the lower figure shows the unthresholded spectrogram with the detected whistles. Common dolphins aggregate in large groups and typically have many overlapping whistles.



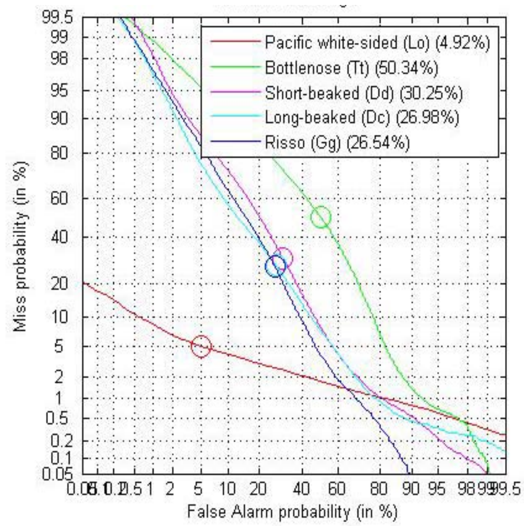


Figure 2 (color online) - Detection error tradeoff curves for a species detection task when the impostor species has not been seen in the training data. A circle shows the point on each curve where the miss and false alarm probabilities are equal with the associated error rate listed in the legend. The left graph shows the performance when Gaussian mixture models are trained for each of the five species except the one being tested and the decision is based on a likelihood ratio of the targeted species to the maximum impostor species is used. The right graph shows performance when a Gaussian mixture model is created by Bayesian adaptation of a universal background model trained from all species except the species being tested (the targeted species and the species associated with impostor calls). Curves that are closer to the origin have better performance.

Experiments from the universal background model are for the moment inconclusive. We ask the question of whether or not a set of 100 consecutive echolocation clicks were produced by a specific species. All feature data from each sighting are randomly assigned to one of three partitions and a three-fold cross validation experiment is run 100 times. Within each fold, the hypothesis is tested against test data from a specific species and one of the other species. For each of the impostor species, no training data from that species is used during model creation. In the context of a baseline Gaussian mixture model system based on our previous work (Roch et al., 2008), this simply means that the model for that species is not used. For the universal background model, a background model is trained using data from species other than the two being tested and a species specific model is created by adapting the background model's means (Reynolds et al., 2000). Figure 2 shows a pair of detection error tradeoff (DET) curves (Martin et al., 1997) which are similar to receiver operating characteristic curves but scale the axes based on normal deviates. The left DET plot shows performance for Gaussian mixture models and the right DET plot is for the universal background model. While the universal background model does improve performance for some species, it is at the expense of poorer performance for others and cannot be said to represent an improved technique at this time.

We hypothesize that some of the difficulty may come from working with a limited number of species. The three species of odontocetes are likely to be insufficient to characterize a general odontocete model. We plan on supplementing this data with additional data to test this hypothesis. Investigation of the problematic classification cases has led us to visualize echolocation clicks through click spectra that are sorted by peak frequency (see Figure 3). Analysis of these plots revealed several trends that had escaped observation by a trained analyst. A couple of our sightings such as the one shown had echo sounder pings that were admitted into the analysis by our click detector. When the spectra are visualized in this manner, many types of anomalies in the data set become easily detectable. In addition to the echo sounders, we observed a number of spectra with very low and very high peak frequencies, some of which appear to be clipped. While this does not affect background model identification any more than our baseline Gaussian mixture models, revisions to our feature extraction algorithm to address this results in over a 15% reduction in error rate in a species identification task on five Southern California odontocetes when compared to our previous method of feature extraction (Roch et al., 2008) on the same data. This reduction is a significant contribution to the research. Results of the species identification task which had a mean error rate of 28% using a similar randomized experiment are shown in Figure 5.

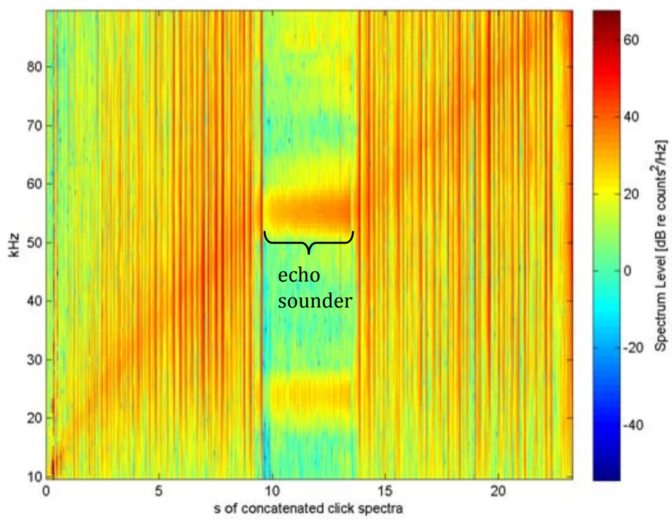


Figure 3 (color online) - Spectra of over 13000 impulsive events recorded in the presence of a November 14, 2004 sighting of long-beaked common dolphins (Delphinus capensis) and sorted by peak frequency. Within each peak frequency, events are sorted by peak energy. Harmonics of an echo sounder becomes easy to detect at 28 and 56 kHz. The effects of D. capensis's orientation with respect to the hydrophone are clearly visible, with presumed on axis echolocation clicks being more to the right although other factors such as slant distance to the hydrophone can have profound effects. This signal variability contributes significantly to making species identification a challenging problem.

he final project sponsored by this work is the random projection tree submanifold learning algorithm (RP-Tree). This project is theoretically the most complex of the three projects and currently the least advanced (as planned in our schedule). We have integrated Dasgupta and Freund's learner into our test framework and implemented tree pruning methods proposed by Quinlan (1993) for his influential C4.5 tree classifier. These additions were completed in the late spring early summer and we have begun to analyze experiments to determine how the model should be refined. Using the randomized cross-validation methods described earlier, we have trained RP-Trees on the same Southern California odontocete classification task. The overall error rate is 37.9%, which is significantly higher than that of the Gaussian mixture model, but for preliminary experiments these results are not unreasonable (histogram not shown due to space constraints). Our current strategy is to examine the misclassifications and determine whether or not the tree pruning method is suitable and to develop an alternative pruning method based upon our observations.

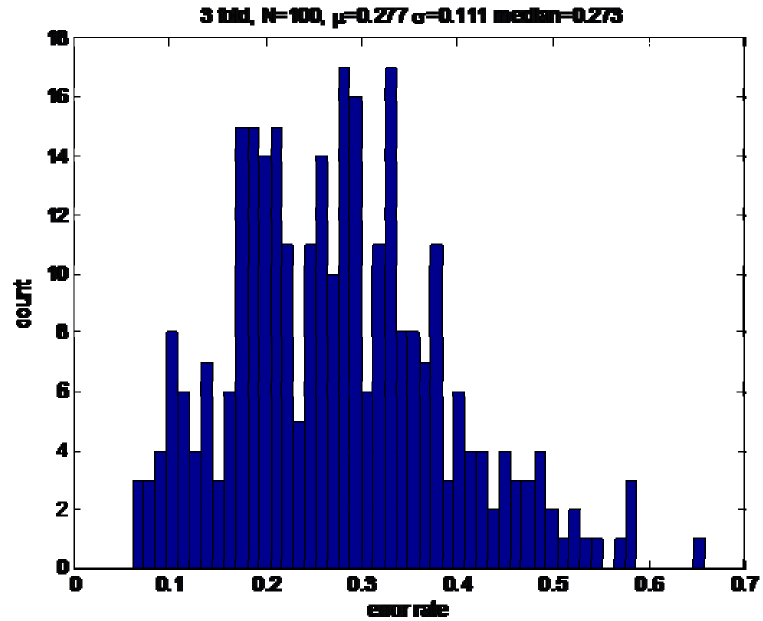


Figure 4 - Error distribution for 100 randomized Gaussian mixture model experiments using our new feature extraction algorithm with three fold cross validation on a species identification task for five Southern California odontocetes: bottlenose dolphin, long and short-beaked common dolphins, Pacific white-sided dolphin and Risso's dolphin. Mean overall error rate: 28%. The previous feature extraction method had a mean error rate of 33%.

IMPACT/APPLICATIONS

This work can be used in passive acoustic monitoring platforms for mitigation and science.

REFERENCES

Dasgupta, S. and Freund, Y. (2007). Random projection trees and low dimensional manifolds. In *Dept. of Computer Science and Engineering*, pp. 29. San Diego: University of California, San Diego.

Dasgupta, S. and Freund, Y. (2008). Random projection trees and low dimensional manifolds. In *Symp. on Theory of Comp. (STOC)*, pp. 10. Victoria (BC), Canada.

Freund, Y., Dasgupta, S., Kabra, M. and Verma, N. (2007). Learning the structure of manifolds using random projections. In *Adv. in Neural Info. Proc. Systems (NIPS)*. Vancouver, B.C., Canada.

Martin, A., Doddington, G., Kamm, T., Ordowski, M. and Przybocki, M. (1997). The DET curve in assessment of detection task performance. In *Eurospeech*, vol. 4, pp. 1895-1898. Rhodes, Greece.

Quinlan, J. R. (1993). C4.5: programs for machine learning. San Mateo: Morgan Kaufmann.

Reynolds, D. A., Quatieri, T. F. and Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. In *Digital Signal Processing*, vol. 10, pp. 19-41.

Roch, M. A., Soldevilla, M. S., Hoenigman, R., Wiggins, S. M. and Hildebrand, J. A. (2008). Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes. *Canadian Acoust.* 36, 41-47.